



09/927,712
45

日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出 願 年 月 日

Date of Application:

2001年 6月14日

出 願 番 号

Application Number:

特願2001-179484

出 願 人

Applicant(s):

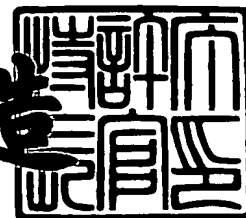
株式会社日立製作所



2001年 8月17日

特許庁長官
Commissioner,
Japan Patent Office

及 川 耕 造



出証番号 出証特2001-3072579

【書類名】 特許願

【整理番号】 K00019721A

【あて先】 特許庁長官

【国際特許分類】 G06F 12/00

【発明者】

 【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所 システム開発研究所内

 【氏名】 藤田 高広

【発明者】

 【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所 システム開発研究所内

 【氏名】 北村 学

【発明者】

 【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所 システム開発研究所内

 【氏名】 木村 光一

【特許出願人】

 【識別番号】 000005108

 【氏名又は名称】 株式会社日立製作所

【代理人】

 【識別番号】 100075096

 【弁理士】

 【氏名又は名称】 作田 康夫

【手数料の表示】

 【予納台帳番号】 013088

 【納付金額】 21,000円

【提出物件の目録】

 【物件名】 明細書 1

 【物件名】 図面 1

特 2 0 0 1 - 1 7 9 4 8 4

【物件名】	要約書	1
【プルーフの要否】	要	

【書類名】 明細書

【発明の名称】 計算機システム

【特許請求の範囲】

【請求項 1】

データを処理するサーバ計算機において、

前記データはディスクに格納されており、前記データを処理するサーバ計算機とは別のサーバ計算機のアクセス手段が、前記ディスクにアクセスできるようにすることで、前記データを処理するサーバ計算機を変更することを特徴とする計算機システム。

【請求項 2】

請求項 1 記載の計算機システムにおいて、

前記データを分割し、分割数分のディスクに格納し、サーバ計算機のアクセス手段が、前記分割数分のディスクにアクセスすることにより、ディスク負荷分散を行うことを特徴とする計算機システム。

【請求項 3】

請求項 1 記載の計算機システムにおいて、

前記データを処理するサーバ計算機に対して、前記データを格納するディスクとの物理的接続に使用しているストレージ装置及びサーバ計算機のポートを変更することにより、ポートの負荷分散を行うことを特徴とする計算機システム。

【請求項 4】

請求項 2 記載の計算機システムにおいて、

前記データを分割し、分割数分のディスクに格納するときに、ストレージ装置のデータコピー手段を使用することを特徴とする計算機システム。

【請求項 5】

データに対する処理を、前記データを分割し、分割したデータを複数のサーバ計算機によって処理するために、前記データに対する処理を前記分割したデータに対する処理に分割する手段を備えた計算機システムにおいて、

前記分割データはそれぞれディスクに格納されており、前記分割データを処理するサーバ計算機とは別のサーバ計算機のアクセス手段が、前記ディスクにア

セスするようにすることで、前記分割データを処理するサーバ計算機を変更することを特徴とする計算機システム。

【請求項 6】

請求項 5 記載の計算機システムにおいて、

分割データを処理するサーバ計算機の変更により、システムの負荷を分散することを特徴とする計算機システム。

【請求項 7】

請求項 5 記載の計算機システムにおいて、

前記分割データをさらに分割し、分割数分のディスクに格納し、サーバ計算機のアクセス手段が、前記分割数分のディスクにアクセスすることにより、ディスク負荷分散を行うことを特徴とする計算機システム。

【請求項 8】

請求項 5 記載の計算機システムにおいて、

前記分割データを処理するサーバ計算機に対して、前記分割データを格納するディスクとの物理的接続に使用しているストレージ装置及びサーバ計算機のポートを変更することにより、ポートの負荷分散を行うことを特徴とする計算機システム。

【請求項 9】

請求項 7 記載の計算機システムにおいて、

前記分割データをさらに分割し、分割数分のディスクに格納するときに、ストレージ装置のデータコピー手段を使用することを特徴とする計算機システム。

【請求項 10】

データを処理する複数のサーバ、

前記データを格納する複数のディスク装置、

前記複数のサーバと前記複数のディスク装置との接続を切り替える手段、

前記データの処理に対する負荷を検出し、前記複数のディスク装置の 1 つにデータの分割を指示し、前記切り替え手段に前記サーバと前記ディスク装置との接続を支持する管理サーバを有することを特徴とする計算機システム。

【発明の詳細な説明】

【 0 0 0 1 】

【発明の属する技術分野】

ストレージエリアネットワーク環境上の分散計算機システムの管理方法に関する。

【 0 0 0 2 】

【従来の技術】

複数の計算機に処理を分散させることによって性能の向上を計る分散型の計算機システムが広く使用されている。このような計算機システムは、クライアントに対し、複数のサーバ計算機によってサービスを提供している。

【 0 0 0 3 】

また、このような計算機システムで使用される計算機は、ストレージ装置と接続するためのポートを複数もち、ストレージ装置にも複数の計算機と接続するためのポートを複数もっている。ストレージ装置内のディスクと計算機は、このポートを介して接続されるが、ディスクと計算機との間の接続があるポートに集中してしまうと、ポートの処理能力を超えてしまうため、ポートが複数ある場合にはディスクと計算機との間の接続を別のポートを介するようにポートの負荷分散を行う。

【 0 0 0 4 】

【発明が解決しようとする課題】

従来の分散型の計算機システムでは、サーバ計算機とストレージ装置は、1対1で接続されおり、サーバ計算機のメンテナンス、またはシステム全体の処理性能向上・負荷分散のために、サーバ計算機及びストレージ装置の追加・除去を伴う計算機システムの物理構成の変更を行う場合、計算機システムのサービスを完全に停止した後、システム管理者が、サーバ計算機とストレージ装置との接続、サーバ計算機のシステム設定の変更を行ったのち、システムのサービス開始を行っており、システム管理者は多くの設定作業を行う必要がある。

また、ポート負荷分散を行う場合にも、計算機システムを停止してのシステム管理者によるシステム再設定が必要となる。

【 0 0 0 5 】

本発明では、分散型の計算機システムにおいて、システムの提供するサービスに対して、サーバ計算機のメンテナンス、システム処理性能の向上・負荷分散のために、サーバ計算機の追加・削除、ディスク負荷分散のためのデータの複数ディスクへの分割を、システムを完全に停止することなく行い、データを処理するサーバ計算機を自動的かつ容易に変更する方法を提供することにより、計算機システムの信頼性の向上およびシステム管理者の負担軽減と設定ミスを低減する。

【 0 0 0 6 】

また、システムを完全に停止することなく、計算機とストレージ装置とを接続するポートの変更を行う手段を提供する。

【 0 0 0 7 】

【課題を解決するための手段】

サーバ計算機とストレージ装置との間に、ネットワークを構築してサーバ計算機とストレージ装置とを物理的に接続する。計算機システムのサービスの処理対象となるあるまとまったデータを分割し、分割したデータ（以後、分割データと呼ぶ）を処理する複数の計算機と分割データを格納するディスクとの間をネットワークで接続し、ある分割データを処理する計算機の負荷が高くなった場合に、計算機が処理する分割データの一部を別の計算機に処理させるために、分割データが格納されているディスクを移動先の計算機に通知し、分割データに関する処理を移行する。このとき、あるまとまったデータに関する処理を提供するサービスの設定に、分割データに関する処理の移行を反映する。

【 0 0 0 8 】

さらに、あるまとまったデータの分割データをさらに分割し、さらに分割されたデータを複数のディスクに分割することによりディスク負荷の分散を行った場合に、さらに分割されたデータを格納するディスクを計算機に通知し、さらに分割されたデータに関する処理を行わせる。このとき、あるまとまったデータに関する処理を提供するサービスの設定に、さらに分割データを分割したごとと分割データを処理する計算機が割り当てられたことを反映する。

【 0 0 0 9 】

また、ユーザに対して、サービスを提供する複数のサーバ計算機をシンボルで

表わし、前記シンボルを画面上で移動させることにより、データを処理するサーバ計算機をサービスから取り除いたり、加えたりするシステム構成の変更を行なうことができるインターフェイスを提供する。

【0010】

【発明の実施の形態】

－第1の実施例－

図1は、本発明を適用する計算機システムの構成図である。計算機システム1は、クライアント計算機10a、10b（総称してクライアント10と呼ぶ）と、FES(Front End Server)計算機100、BES(Back End Server)計算機200、これらサーバ計算機にファイバチャネルスイッチ40を介して接続されるストレージ装置50a、50b（総称してストレージ装置50と呼ぶ）とネットワーク60から構成される。

【0011】

ストレージ装置50は、ディスク510、サーバ計算機30と接続するストレージポート53から構成される。本実施例におけるディスク510は、論理的なデバイスであり、複数の物理ディスクから構成されているが、実施例の説明には関係ないため省略する。ただし、ディスク510は、論理的なデバイスに限らない。

【0012】

ストレージポート53は、接続されるサーバ計算機30がUNIXなどのオープンシステムであればSCSI(Small Computer System Interface)のインタフェース、あるいはサーバ計算機がメインフレームであればESCON(Enterprise System Connection)などのチャネルインタフェースであり、それぞれのストレージポート53が同一のインタフェースであっても異なるものが混在していてもかまわない。本実施例では、全てのインタフェースがファイバチャネルインタフェースであるとする。

【0013】

ファイバチャネルスイッチ40は、サーバポート32とストレージ装置50のストレージポート53との間の接続を制限するゾーニング機能をもっており、ネ

ットワークインタフェース32を使って、ゾーニングを設定できる。同一のゾーン内ではサーバポート32からストレージポート53に対してアクセスできるが、サーバポート32とストレージポート53とが、それぞれ異なるゾーンに属しているときはアクセスできない。このゾーニングの機能によって、アクセスを制限できる。ゾーニングは必ずしも必要でないが、ゾーニングによってサーバ30からストレージ装置50への不用意なアクセスを防ぐことができる。今後、ポート間の接続とは、ポートが同一のゾーン内にあることを指す。

【0014】

ストレージ装置50は、ディスク510のデータを別のディスク510にコピーする機能をもっている。この機能は例えば、米国特許5,845,295号にて開示されている。

【0015】

また、ストレージ装置50は、ディスク510に対して、少なくとも計算機システム1内で一意であることが保証されるIDをつける。例えば、ストレージ装置のベンダ名とストレージ装置のシリアル番号とディスクの内部識別子を使用してストレージIDを生成する。

【0016】

計算機システム1は、少なくとも1つ以上のサービスを1つ以上のクライアント計算機に対して提供するものである。本計算機システムで提供されるサービスに対するクライアントからの要求は、ある意味をもつデータに対する処理に関するものである。サービスは、1つのFESと少なくとも1つ以上のBESによって提供される。ある意味をもつデータはすくなくとも1つ以上に分割されて、ディスクに格納される。FESは、クライアント計算機からの処理要求を分割されたデータに対する処理に分解する。分割されたデータは、ディスクに格納されており、ディスクに対してアクセスし、格納されている分割データを処理するように指定されたBESが処理を行う。

【0017】

例えば、このようなサービスとして、ファイルサーバがある。FESは、クライアントに対してある意味があるデータとしてディレクトリ構造（仮想ディレク

トリ)を提供する。データは、ある上位のディレクトリごとに分割されディスクに格納される。クライアントからの、仮想ディレクトリ中のファイルに対するリード要求は、そのファイルへのパス名からディレクトリごとに分割された分割データに対する処理に分解されて、この分割データを格納したディスクに対してアクセスし、分割データを処理するように指定されているBESに処理要求が送られる。要求を受けたBESは、ディスクからデータを読み出して、指定されたファイルのデータをクライアントに渡す。このようなサービスとして、他にWebサーバ、データベースシステムがある。

【0018】

図2は、FES100のブロック構成図である。FES100は、クライアント処理分割手段110、データ処理BES変更手段120、ディスク分割手段130を備え、FESシステム管理情報150により構成されている。

【0019】

クライアント処理分割手段110は、クライアントからの要求を、FESシステム管理情報150を参照して、分割データに対する処理に分割し、分割データを担当するBES200に割り振る。

【0020】

データ処理BES変更手段120は、クライアント処理分割手段120が、分割データを処理するBES200を割り振る方法を変更する。つまり、ある分割データがあるBESが処理するようにクライアント処理分割手段120が割り振っているとき、データ処理BES変更手段120による変更により、前記ある分割データは、前記あるBESとは別のBESによって処理されるようになる。

【0021】

ディスク分割手段130は、クライアント処理分割手段120が、クライアントからの要求を、分割する分割データをさらに分割する。つまり、ある分割データがあるときに、これをさらに2つ以上の分割データに分割し、これらさらに分割された分割データは、適切なBESによって処理されるようにする。

【0022】

ディスク負荷収集手段140は、BES200のディスク負荷取得手段230

が取得するディスクごとのディスク負荷を収集し、FESシステム管理情報150のディスク負荷のエントリを更新するものであり、ある一定時間ごとに実行される。

【0023】

FESシステム管理情報150は、計算機システムが提供するサービスの各FESが管理する情報である。このFESシステム管理情報150は、図3に示すように、分割データ識別子、分割データが格納されているディスクのディスクID、分割データを処理するため前記分割データを格納する前記ディスクにアクセスするBESと前記ディスク負荷を管理する。クライアント処理分割手段110は、この情報をもとに、クライアントからの要求の分割とBESへの割り振りを行う。データ処理BES変更手段120は、この情報を変更して、分割データを処理するBESを変更する。ディスク分割手段により、分割データがさらに分割された場合には、この情報にエントリが追加される。

【0024】

図4は、BES200のブロック構成図である。BES200は、データ処理手段210、ディスクアクセス手段220、ディスク負荷取得手段230、アクセスディスク確認手段240から構成される。

【0025】

データ処理手段210は、ディスクアクセス手段220からアクセスできるディスクに格納されたデータを処理する。具体的には、ディスクに格納されたデータを読み出しプロトコルに従いファイルを提供する処理、ディスクに格納されたデータの検索・更新を行うデータベース処理、ディスクに格納されたデータを読み出しプロトコルに従いデータを提供する処理である。

【0026】

ディスクアクセス手段220は、ストレージ装置50のディスクに対してアクセスを行う。

ディスク負荷取得手段230は、BESのディスクアクセス手段220がアクセスするディスクの負荷を取得するものであり、たとえばある一定時間のI/O数を取得し、ディスクが処理できる最大I/O数に対する%を取得し、FES1

00のディスク負荷収集手段140に対して、ディスク負荷情報を提供する。

【0027】

アクセスディスク確認手段240は、少なくともシステムで一意であることが保証されているディスクIDを調べることにより、ディスクアクセス手段220によって、あるディスクに対してアクセスできるかを確認するものである。

【0028】

図5は、ストレージ装置50のブロック構成図である。ストレージ装置50は、データ格納手段であるディスク510、データコピー手段550、ポートマッピング変更手段560を備える。

【0029】

ディスク510は、BES200のディスクアクセス手段220によってアクセスされるものであり、少なくとも計算機システム1で一意であるディスクIDにより、一意に識別される。図5には、ディスク510は、一つしか書いていないが、ストレージ装置510には、一般に複数のディスク510が存在している。

【0030】

データコピー手段550は、ディスク510が格納するデータをこのディスク510とは別のディスク510にコピーするものであり、コピー先のディスク510は、コピー元のディスクと同じストレージ装置50である必要はなく、別のストレージ装置であってもよい。

【0031】

ストレージ装置50は、BES200からディスク510に対するアクセスを受け付けるポートを複数備えており、ディスク510をどのポートを介してBES200と接続するかを決めて変更できる。ポートマッピング変更手段560は、この変更を行なう。

【0032】

図6は、システム管理サーバ400のブロック構成図である。システム管理サーバ400は、GUI410、FESシステム管理情報取得手段420、BES負荷監視手段430、ディスク負荷監視手段440、ポート負荷監視手段450

、B E S 追加処理 4 6 0、B E S 削除処理 4 7 0、サーバ情報 4 8 0、ディスク情報 4 8 5、ストレージポート情報 4 9 0、ポート接続情報 4 9 5 から構成される。

【 0 0 3 3 】

G U I 4 1 0 は、ユーザに対して、計算機システム 1 によって提供されるファイルサーバ、データベースサーバ、W E B サーバなどのサービスを容易に管理するためのインターフェイスを提供するものであり、図 7 のように、F E S、B E S、ディスクをシンボルで表し、B E S がディスクを担当している場合はディスクと B E S とを線で接続し、システムで提供されるサービスごとに矩形が用意されサービスを提供している F E S と B E S とディスクとを同一の矩形内に表示し、使用していない B E S およびディスクは未使用を表す矩形内に表示する。

【 0 0 3 4 】

ユーザが、サービスを表す矩形内の B E S のシンボルを未使用矩形に移動すると、B E S 削除処理 4 7 0 が実行され、未使用矩形に移動された B E S が担当していた分割データは、サービスを提供している他の B E S に振り分けられる。これにより、サーバメンテナンス時に、G U I による簡単な操作だけで、サービスを停止することなく B E S をサービスから切り離すことができる。サーバメンテナンス終了時には、ユーザは、G U I 上で未使用領域の B E S を、サービスを表す矩形に戻す操作により、サーバ追加処理 4 6 0 が実行され、サービスの全体性を元の状態に戻すことができる。

【 0 0 3 5 】

F E S システム管理情報取得手段 4 2 0 は、F E S 1 0 0 の F E S システム管理情報 1 4 0 を取得するものである。

【 0 0 3 6 】

B E S 負荷監視手段 4 3 0 は、B E S 2 0 0 の負荷を取得し、サーバ情報 4 8 0 を更新する手段であり、負荷の大きさがあるしきい値をこえるなどにより、B E S 2 0 0 の過負荷を検出すると、サーバ負荷分散のため、データ処理 B E S 変更処理 1 2 0 を行なうものであり、ある一定時間ごとに実行される。

【 0 0 3 7 】

ディスク負荷監視手段440は、FESシステム管理情報取得手段420を使用して、FES100のFESシステム管理情報140を取得し、ディスク負荷の大きさがあるしきい値をこえるなどにより、ディスクの過負荷を検出すると、ディスク負荷分散のため、ディスク分割処理130を行なうものであり、ある一定時間ごとに実行される。

【0038】

ポート負荷監視処理450は、ストレージポートの負荷をストレージ装置から取得し、ストレージポート情報490のポート負荷を更新する手段であり、負荷の大きさがあるしきい値を超えるなどにより、ポートの過負荷を検出すると、ポート負荷分散のため、サーバとの接続のためにディスクが介しているストレージ装置のポートを変更するものであり、ある一定時間ごとに実行される。

【0039】

BES追加処理460は、GUI410上でユーザが計算機システム1が提供するサービスにBES200を追加する操作を行なったときに実行される処理であり、新規追加するBES200のディスクアクセス手段230によって、ある分割データを格納するディスクに対してアクセスできるようにし、追加先のサービスの分割データを処理できるようにする。

【0040】

BES削除処理470は、GUI410上でユーザが計算機システム1が提供するサービスからBES200を削除する操作を行なったときに実行される処理であり、削除するBES200が担当していた分割データを格納するディスクに対して、同じサービスを提供する他のBES200のディスクアクセス手段230がアクセスできるようにすることで、削除するBESが処理を担当していた分割データに関する処理を、別のBESに移行するものである。

【0041】

サーバ情報480は、計算機システム1で使用されるFES100およびBES200に関するものであり、図8に示すように、BES名と現在所属しているFES名と所属できるFESのリストとBESのサーバ負荷からなるテーブルとFES名とFESによって提供されるサービスのシステム名とからなるテーブル

からなる。

【0042】

ディスク情報485は、計算機システム1にあるディスク510に関するものであり、図9に示すように、ディスクIDと現在の使用状況とディスクが含まれているストレージ装置とからなるテーブルである。ディスク情報485は、ディスク分割するときに未使用のディスクがあるかを調べる時に使用される。

【0043】

ストレージポート情報490は、ストレージ装置のポートに関するものであり、図10に示すように、ストレージポートのポート名とストレージ装置のIDとディスク数とポート負荷とからなるテーブルである。ストレージポート名は、ストレージ装置のポートのWorld Wide Name（以下WWNと略す）であり、ストレージ装置IDは、システム内で一意なストレージ装置の識別子である。ディスク数は、前記ストレージポートを介してBES200と接続しているディスク510の数であり、ポート負荷は前記ストレージポートの負荷である。

【0044】

ポート接続情報495は、BES200とディスク510との間の接続に使用しているポートに関するものであり、図11に示すように、サーバ名とサーバポート名とストレージポート名とディスクIDとからなるテーブルである。サーバ名は、ディスク510に対してアクセスするBES200の名前であり、サーバポート名は、前記アクセスに使用するサーバポートのWWNである。ストレージポート名は、前記アクセスに使用するストレージ装置50のポートのWWNであり、ストレージ装置IDは、前記ストレージ装置50の識別子である。

【0045】

次に、この実施例における処理の流れを図をもって説明する。

【0046】

GUI410にて、BESを削除する操作となるBESのシンボルを、サービスを表す矩形内から未使用矩形に移動したときの、BES削除処理470を図12で説明する。

【0047】

ステップ6010で、サーバ情報480から削除指定されたBESが属するFESを特定し、特定したFESからFESシステム管理情報150を取得する。

【0048】

ステップ6020で、FESシステム管理情報150から削除指定されたBESが担当している全分割データと分割データを格納するディスクのディスクIDとを取得する。

【0049】

ステップ6030で、サーバ情報480から削除指定されたBESと同じFESに属するBESの中から分割データを引き継ぐ移動先BESを決定する。

【0050】

ステップ6040で、移動先BESを決定できなかったときは、ステップ6100でユーザに削除不可を報告して終了する。決定できたときは、ステップ6050にすすむ。

【0051】

ステップ6050で、削除指定されたBESを移動元BES、移動元BESが担当していた分割データを移動分割データ、ステップ6030で決定した移動先BESとを指定して、移動元BESが属するFESのデータ処理BES変更手段120を実行する。

【0052】

ステップ6060で、BESを変更できなかったときは、ステップ6100でユーザに削除不可を報告して処理を終了する。変更できたときは、ステップ6070で削除指定されたBESのサーバ情報480のエントリを未使用に変更する。

【0053】

この処理は、あるBESが故障したときにシステムが提供するサービスを継続するために、データ処理を別のBESに引き継ぐときにも実行することができる。

【0054】

GUI410にて、BESを追加する操作となるBESのシンボルを未使用を

表す矩形内からサービスを表す矩形に移動したときの、B E S 追加処理 4 6 0 を図 1 3 で説明する。

【 0 0 5 5 】

ステップ 7 0 1 0 で、新規追加 B E S は、追加先サービスの F E S に属することができるかを、サーバ情報 4 8 0 で調べる。所属できないときは、ステップ 7 1 0 0 でユーザに追加不可を報告して終了する。所属できるときは、ステップ 7 0 2 0 に進む。

【 0 0 5 6 】

ステップ 7 0 2 0 で、追加先の F E S から F E S システム管理情報 1 5 0 を取得する。

【 0 0 5 7 】

ステップ 7 0 3 0 で、F E S システム管理情報 1 5 0 から追加 B E S に割り当てる移動分割データを決定する。

【 0 0 5 8 】

ステップ 7 0 4 0 で、移動分割データを決定できなかったときは、ステップ 7 1 0 0 でユーザに追加不可を報告して終了する。決定できたときは、ステップ 7 0 5 0 に進む。

【 0 0 5 9 】

ステップ 7 0 5 0 で、移動分割データを担当していた B E S を移動元 B E S、移動分割データ、追加 B E S を移動先 B E S として F E S のデータ処理 B E S 変更手段 1 2 0 を実行する。

【 0 0 6 0 】

ステップ 7 0 6 0 で、B E S を変更できなかったときは、ステップ 7 1 0 0 でユーザに追加不可を報告して終了する。決定できたときは、ステップ 7 0 7 0 に進む。

【 0 0 6 1 】

ステップ 7 0 7 0 で、サーバ情報 4 8 0 の追加 B E S のエントリの所属 F E S を B E S 追加したサービスの F E S に変更する。

【 0 0 6 2 】

この処理は、メンテナンス終了時にメンテナンスを行ったBESをサービスに戻るとき以外にも、サービス全体の負荷が高くなった場合に、BESを追加してサービスの全体性能を向上させるときにも実行される。

【0063】

次に、BES削除処理470、BES460で実行されるデータ処理BES変更処理120を、図14を使って説明する。

【0064】

ステップ2010で、FESのクライアント処理分割手段110に対して、サービス一時停止を指示する。これにより、クライアント計算機からの処理の分割とBESに対する処理の割り振りを一時中断する。

【0065】

ステップ2020で、分割データに対する処理の移行先BESに対して、分割データを格納するディスクのディスクIDを渡し、ディスクアクセス可確認を指示する。この指示を受けたBESは、アクセスディスク確認手段240により、指定されたディスクIDを持つディスク510に対してアクセスできるかを確認する。

【0066】

ステップ2030で、移行先BESが移行分割データを格納するディスクに対してアクセスできるときは、ステップ2040に進み、できなかったときはステップ2100に進み、変更不可を報告し処理を終了する。

【0067】

ステップ2040で、移行元BESのデータ処理手段210に対して、移行分割データに対するデータ処理の終了を指示する。

【0068】

ステップ2050で、移行先BESのデータ処理手段210に対して、移行分割データに対するデータ処理の開始を指示する。

【0069】

ステップ2060で、FESシステム管理情報150の移行分割データのエントリの移行元BESを移行先BESに変更し、以後の分割データに関する処理が

移行先 B E S に割り振られるようにする。

【 0 0 7 0 】

ステップ 2 0 7 0 で、F E S のクライアント処理分割手段 1 1 0 に対して、サービスの開始を指示する。

【 0 0 7 1 】

この処理は、ある分割データを処理する B E S を変更する処理であり、B E S の負荷を管理する B E S 負荷監視手段 4 3 0 から実行される。また、G U I 4 1 0 上で、B E S を表わすシンボルとディスクを表わすシンボルとを結ぶ線の B E S 側の端を、別の B E S に移動する操作をしたときにも実行される。

【 0 0 7 2 】

図 1 5 は、B E S 2 0 0 のアクセスディスク確認処理 2 4 0 である。アクセスディスク確認処理 2 4 0 は、ディスクを識別するディスク I D を受け取り、前記ディスク I D を持つディスクにアクセスできるかを確認する処理である。

【 0 0 7 3 】

ステップ 4 0 1 0 でディスク I D を持つディスク 5 1 0 に対してアクセスできる時は、ステップ 4 0 2 0 に進みアクセス可を報告する。アクセスできない時は、ステップ 4 1 0 0 に進みアクセス負荷を報告する。ディスクの確認は、S C S I のコマンドである *I n q u i r y* を発行して得られるシリアル番号をディスク I D と比較することで行うこともできるし、ディスク I D をディスクの特定の領域に書き込み、これを読み出すことで行うこともできる。

【 0 0 7 4 】

次に、データ処理を行っているサーバ計算機の負荷を監視し、負荷の高いサーバが見つかった場合には、負荷の高いサーバが処理するデータを別のサーバ計算機が処理するように変更することで、負荷分散を行う B E S 負荷監視処理 4 3 0 を図 1 6 を使って説明する。本処理は、システム管理サーバ 4 0 0 において定期的に実行される。

【 0 0 7 5 】

ステップ 5 0 1 0 でサーバ情報 4 8 0 にエントリのある B E S に対して負荷を問い合わせる。

【0076】

ステップ5020で、サーバ情報480のサーバ負荷を更新する。

【0077】

ステップ5030で、過負荷のサーバが無い場合は、処理を終了する。過負荷のサーバを検出した場合には、ステップ5100へ進む。

【0078】

ステップ5100で、過負荷を検出したBESが所属するFESをサーバ情報480から取得して、FESからFESシステム管理情報150を取得して、過負荷を検出したBESがデータ処理BESとなっているエントリを取り出し、この中から移動するデータ（ディスク）を決定する。決定する方法としては、たとえば最もディスク負荷の大きい分割データを移動データとすればよい。

【0079】

ステップ5110で、サーバ情報480から過負荷のBESと同じFESに属するBESを取り出し、この中からデータ移動先となるBESを決定する。決定する方法としては、たとえば最もサーバ負荷の低いBESを移動先BESとすればよい。

【0080】

ステップ5120で、移動データと移動先BESとが決定できた場合には、ステップ5130に進み、いずれか一方でも決定できなかったときは、ステップ5200に進む。

【0081】

ステップ5130で、過負荷を検出したBESを移動元BES、ステップ5100で決定した移動データ、ステップ5110で決定した移動先BESを指定して、データ処理BES変更処理120を実行する。これにより、移動データを処理するBESは、移動元BESから移動先BESに変わり、移動元BESの負荷が低減される。

【0082】

ステップ5140で、変更できた時は終了し、出来なかったときはステップ5200に進む。

【0083】

ステップ5200で、ユーザに対して、BESの過負荷とディスク移動ができないことを報告して終了する。

【0084】

上記BES負荷監視処理430は、サーバ計算機の負荷分散を行なう処理であるが、サーバ計算機がアクセスするディスク510のI/Oがシステムのボトルネックとなる場合がある。これを防ぐため、システム管理サーバ400で、ディスク負荷監視処理440を定期的に実行する。ディスク負荷監視処理440を、図17を使用して説明する。

【0085】

ステップ8010で、サーバ情報にエントリのあるFESからFESシステム管理情報150を取得する。

【0086】

ステップ8020で、過負荷ディスクがある場合には、ステップ8030に進み、無い場合は終了する。ディスク過負荷の判断は、たとえばFESシステム管理情報150のディスクI/Oがある閾値を超えた場合に過負荷と判断する。

【0087】

ステップ8030で、過負荷ディスクのディスク分割を過負荷ディスクが属するFESのディスク分割手段130を使って実行する。

【0088】

ステップ8040で、ディスク分割できた場合には、処理を終了し、ディスク分割できなかった場合には、ステップ8100に進む。

【0089】

ステップ8100で、ディスク分割できなかった場合に、ユーザに対して、ディスクの過負荷とディスク分割ができないことを報告する。

【0090】

本処理では、データ処理を行っているBESがアクセスするディスクの負荷を監視し、負荷の高いディスクが見つかった場合には、負荷の高いディスクに格納されるデータをさらに分割して別のディスクに格納し、BESがこれらさらに分

割された分割データを格納するディスクにアクセスできるようにすることで、システムの提供するサービスを停止することなく、ディスク負荷の分散を行うものである。

【0091】

ディスク負荷監視処理440から呼び出されるディスク分割処理130を、図18を使用して説明する。

【0092】

ステップ3010で、分割対象データの分割数と分割データIDとを決定する。

【0093】

ステップ3020で、ディスク情報485を参照して分割数分のディスクがある事を確認する。

【0094】

ステップ3030で、分割数分のディスクがある場合には、ステップ3040に進み、ない場合には、ステップ3200で分割不可を報告する。

【0095】

ステップ3040で、ディスク情報485の、ステップ3020で確保したディスクに対応するエントリの使用状況を使用に変更する。

【0096】

ステップ3050で、ストレージ装置50のデータコピー手段550に対して、分割対象ディスクからステップ3020で確保したディスクに対してディスクコピーを指示する。

【0097】

ステップ3060で、分割対象データを担当しているBES（以後、本処理の説明では対象BESとよぶ）が所属しているFESのクライアント処理分割手段110にサービス一時中断を指示する。

【0098】

ステップ3070で、分割対象データを担当しているBESに、ステップ3020で確保したディスク510に対するアクセスディスク確認処理240の実行

を指示する。

【0099】

ステップ3080で、アクセスできる時は、ステップ3090に進む。アクセスできない時は、ステップ3300で、ステップ3020で確保したディスクのディスク情報485のエントリを未使用に戻し、ステップ3310で分割不可を報告し、ステップ3140に進む。

【0100】

ステップ3090で、対象BESに対して、分割対象データに対する処理の終了を指示する。

【0101】

ステップ3110で、対象BESが所属しているFESのFESシステム管理情報150の分割対象データのエントリを削除する。

【0102】

ステップ3120で、対象BESに分割後のデータに対するデータ処理の開始を指示する。

【0103】

ステップ3130で、対象BESが属するFESのFESシステム管理情報150にデータ分割後の分割データID、ディスクID、対象BESからなるエントリを追加する。

【0104】

ステップ3140で、対象BESが属するFESのクライアント処理分割手段110にサービスの再開を指示する。

【0105】

本処理は、ある分割データを少なくとも2つ以上にさらに分割して、前記さらに分割された分割データを格納するディスクを用意し、前記ディスクに対してBESがアクセスできることを確認し、システム運用を継続する処理である。この処理は、ディスク負荷監視手段440の他に、GUI410にて、ある分割データを格納するディスクを選択し、メニューなどからデータ分割を選んだ場合にも実行される。

【0106】

BES 負荷監視処理 430 により BES の負荷が分散され、ディスク負荷監視処理 440 によりディスクの負荷が分散される。しかし、BES 200 は、ポートを介してディスク 510 にアクセスするため、ポートの処理性能がシステムのボトルネックとなることがある。

【0107】

ストレージ装置に複数のポートがある場合に、ポートの負荷を分散させるのが、ストレージポート負荷監視処理 450 であり、図 19 を使用して説明する。

【0108】

ステップ 9010 で、ストレージポートの I/O 負荷をストレージ装置 50 から取得する。

【0109】

ステップ 9020 で、ストレージポート情報 490 のポート負荷を更新する。

【0110】

ステップ 9030 で、過負荷ストレージポートがある場合には、ステップ 9040 に進み、過負荷ストレージポートがない場合には、処理を終了する。

【0111】

ステップ 9040 で、過負荷ストレージポートがある場合には、ポート接続情報 495 から、過負荷ストレージポートを介して接続しているディスクを取り出し、ポートを変更するディスクを決定する。

【0112】

ステップ 9050 で、ストレージポート情報 490 を参照して、過負荷ポートと同一ストレージ装置内に負荷が高くないポートがある場合には、ステップ 9060 にすすみ、同一ストレージ装置内に負荷が高くないポートが無い場合には、ステップ 9200 に進む。

【0113】

ステップ 9200 で、同一ストレージ装置内に負荷が高くないポートが無い場合には、別ストレージ装置に負荷が高くないポートと未使用のディスクがあるかをストレージポート情報 490 とディスク情報 485 とから決定する。決定でき

なかったときは、処理を終了する。決定できたときは、ステップ 9 2 1 0 に進む。

【 0 1 1 4 】

ステップ 9 2 1 0 で、ステップ 9 2 0 0 で確保したディスクに対応するディスク情報 4 8 5 のエントリを使用に変更する。

【 0 1 1 5 】

ステップ 9 2 2 0 で、ストレージ装置に対して、ポートを変更元のディスクからポート変更先のディスクに対してデータコピーを指示する。

【 0 1 1 6 】

ステップ 9 0 6 0 で、ポート接続情報 4 9 5 からポート変更ディスクを使用している B E S を取り出し、この B E S が属する F E S をサーバ情報 4 8 0 から取得した後、取得した F E S のクライアント処理分割手段 1 1 0 にサービス一時中断を指示する。

【 0 1 1 7 】

ステップ 9 0 7 0 で、ポート変更ディスクを担当している B E S にポート変更前のディスクが格納する分割データに対するデータ処理の終了を指示する。

【 0 1 1 8 】

ステップ 9 0 8 0 で、ストレージ装置のポートマッピング変更手段 5 6 0 にポート変更を指示する。

【 0 1 1 9 】

ステップ 9 0 9 0 で、ストレージポート情報 4 9 0 とポート接続情報 4 9 5 とを変更する。

【 0 1 2 0 】

ステップ 9 1 0 0 で、ポート変更ディスクを担当している B E S にポート変更後のディスクに対するディスクアクセス可確認を指示する。

【 0 1 2 1 】

ステップ 9 1 1 0 で、アクセス可のときは、ステップ 9 1 2 0 に進み、アクセス不可のときは、ステップ 9 3 0 0 に進む。

【 0 1 2 2 】

ステップ 9 1 2 0 で、ステップ 9 1 0 0 でアクセスディスク確認処理 2 4 0 を実行した B E S に対して、ポート変更後のディスクに格納される分割データに対するデータ処理の開始を指示する。

【 0 1 2 3 】

ステップ 9 1 3 0 で、ステップ 9 0 6 0 と同じ F E S のクライアント処理分割手段 1 1 0 にサービス再開を指示する。

【 0 1 2 4 】

本処理は、システム管理サーバ 4 0 0 において定期的に実行される。また、本処理では、ストレージポートのポート負荷分散を行なったが、ストレージポート情報 4 9 0 のように、サーバポート情報を管理することで、サーバポートの負荷分散を行なうこともできる。

【 0 1 2 5 】

【発明の効果】

分散計算機システムにおいて、システム全体を停止することなく、サーバ計算機の追加、ストレージ装置の追加による計算機システムの拡張を行うことができる。

【 0 1 2 6 】

サーバ計算機の故障したときに、故障したサーバ計算機に接続していたディスクを別のサーバ計算機に接続を変更することで、計算機システムの動作を継続することができ、システムの耐障害性を向上させることができる。

【 0 1 2 7 】

ディスク負荷分散のために、ディスクに格納されるデータを複数に分割し、同時に分割したデータが処理されるように、システム設定を自動的に変更することで、システムを停止することなく、システム管理者は容易にディスク負荷の分散を実行できる。

【 0 1 2 8 】

ポート負荷の分散のために、サーバとディスクとの間の接続で使用するポートを変更したときに、ポートを変更したディスクに対するアクセスを確認し、同時に設定を自動的に変更することで、システムの変更を容易にし、システム管理者

の操作ミスによるシステムの停止を防ぐことができる。

【 0 1 2 9 】

サーバとディスクとの接続に使用するポートを自動的に変更することによって、サーバとディスクとの間のスループットを維持することができる。

【図面の簡単な説明】

【図 1】

本発明を適用する計算機システムの構成図である。

【図 2】

F E S のブロック構成図である。

【図 3】

F E S が管理する F E S システム管理情報である。

【図 4】

B E S のブロック構成図である。

【図 5】

ストレージ装置のブロック構成図である。

【図 6】

システム管理サーバのブロック構成図である。

【図 7】

システム管理サーバが提供する G U I である。

【図 8】

システム管理サーバが管理するサーバ情報である。

【図 9】

システム管理サーバが管理するディスク情報である。

【図 1 0】

システム管理サーバが管理するストレージポート情報である。

【図 1 1】

システム管理サーバが管理するノード間接続情報である。

【図 1 2】

システム管理サーバの B E S 削除処理の処理フローである。

【図 1 3】

システム管理サーバの B E S 追加処理の処理フローである。

【図 1 4】

F E S のデータ処理 B E S 変更処理の処理フローである。

【図 1 5】

B E S のアクセスディスク確認処理の処理フローである。

【図 1 6】

システム管理サーバの B E S 負荷監視処理の処理フローである。

【図 1 7】

システム管理サーバのディスク負荷監視処理の処理フローである。

【図 1 8】

F E S のディスク分割処理の処理フローである。

【図 1 9】

システム管理サーバのポート負荷監視処理の処理フローである。

【符号の説明】

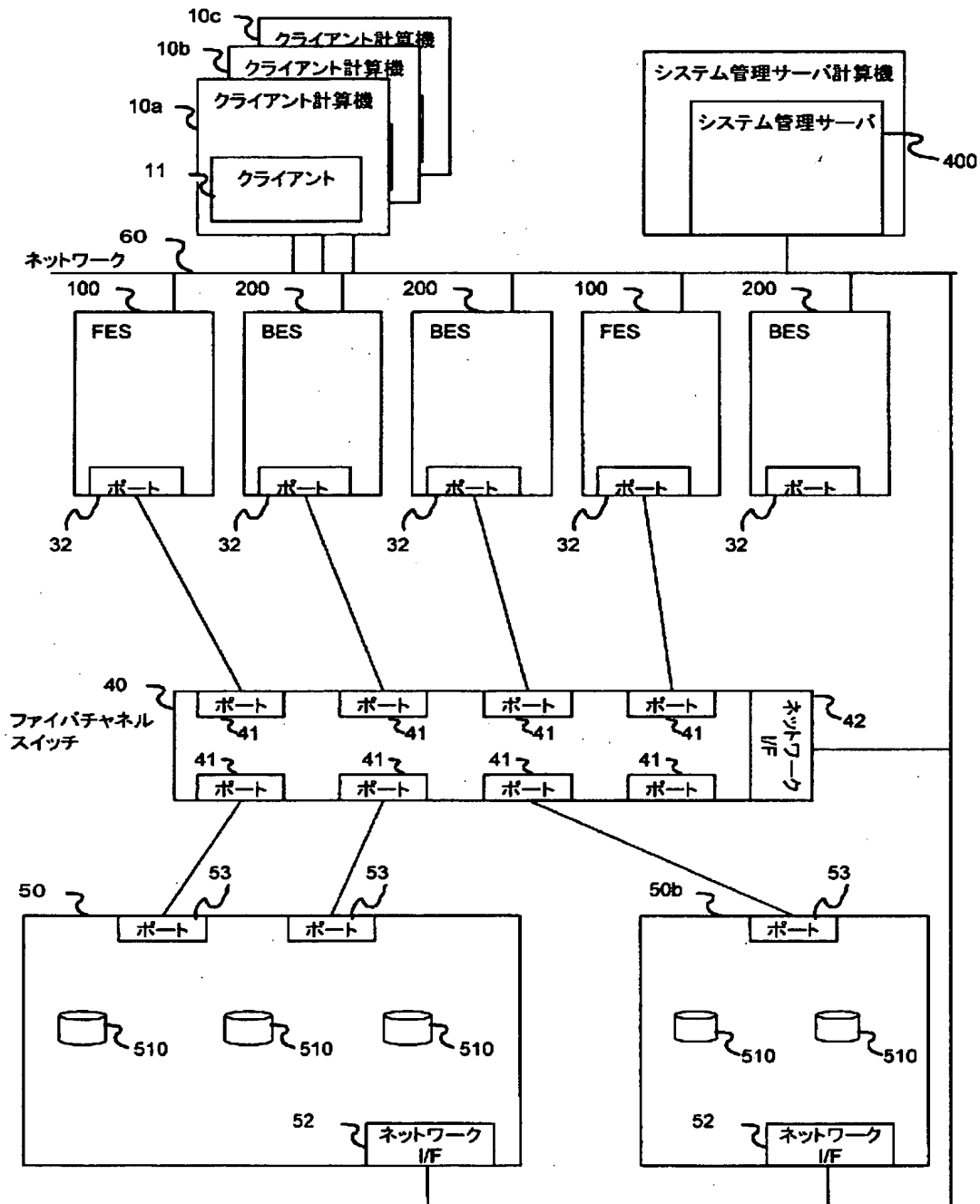
1 … 計算機システム、 1 0 … クライアント計算機、 2 0 … 管理計算機、 3 2 … サーバポート、 4 0 … ファイバチャネルスイッチ、 5 0 … ストレージ装置、 5 1 0 … ディスク、 1 0 0 … F r o n t E n d S e r v e r (F E S) 計算機、 2 0 0 … B a c k E n d S e r v e r (B E S) 計算機

【書類名】 図面

【図 1】

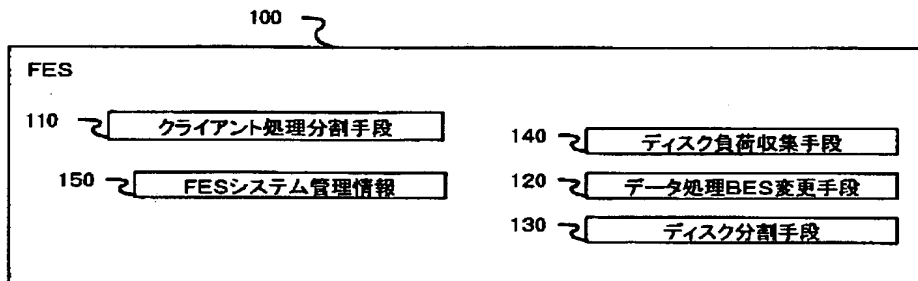
図1

1



【図 2】

図2



【図 3】

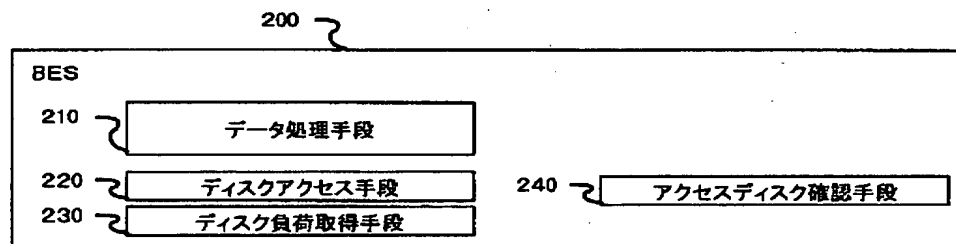
図3

150

分割データ識別子	ディスクID	データ処理BES	ディスク／O
Data1	DID1	BES1	20%
Data2	DID2	BES1	20%
Data3	DID3	BES2	20%

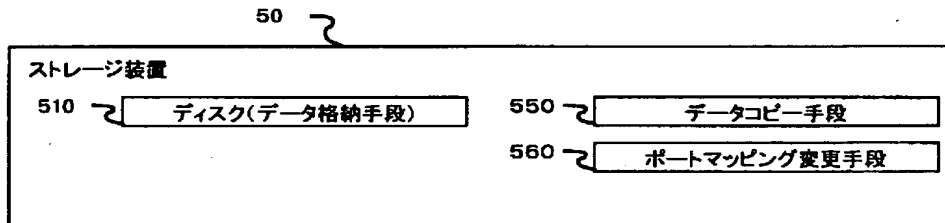
【図 4】

図4



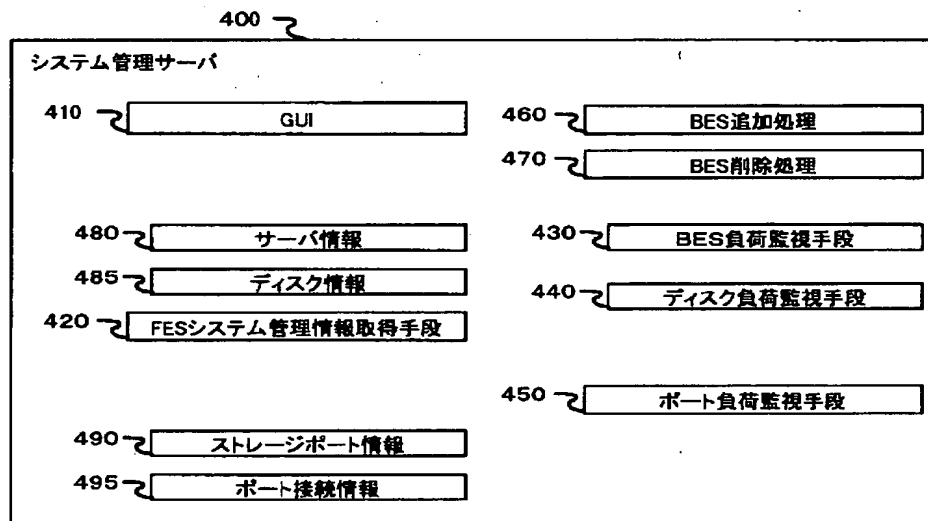
【図 5】

図5



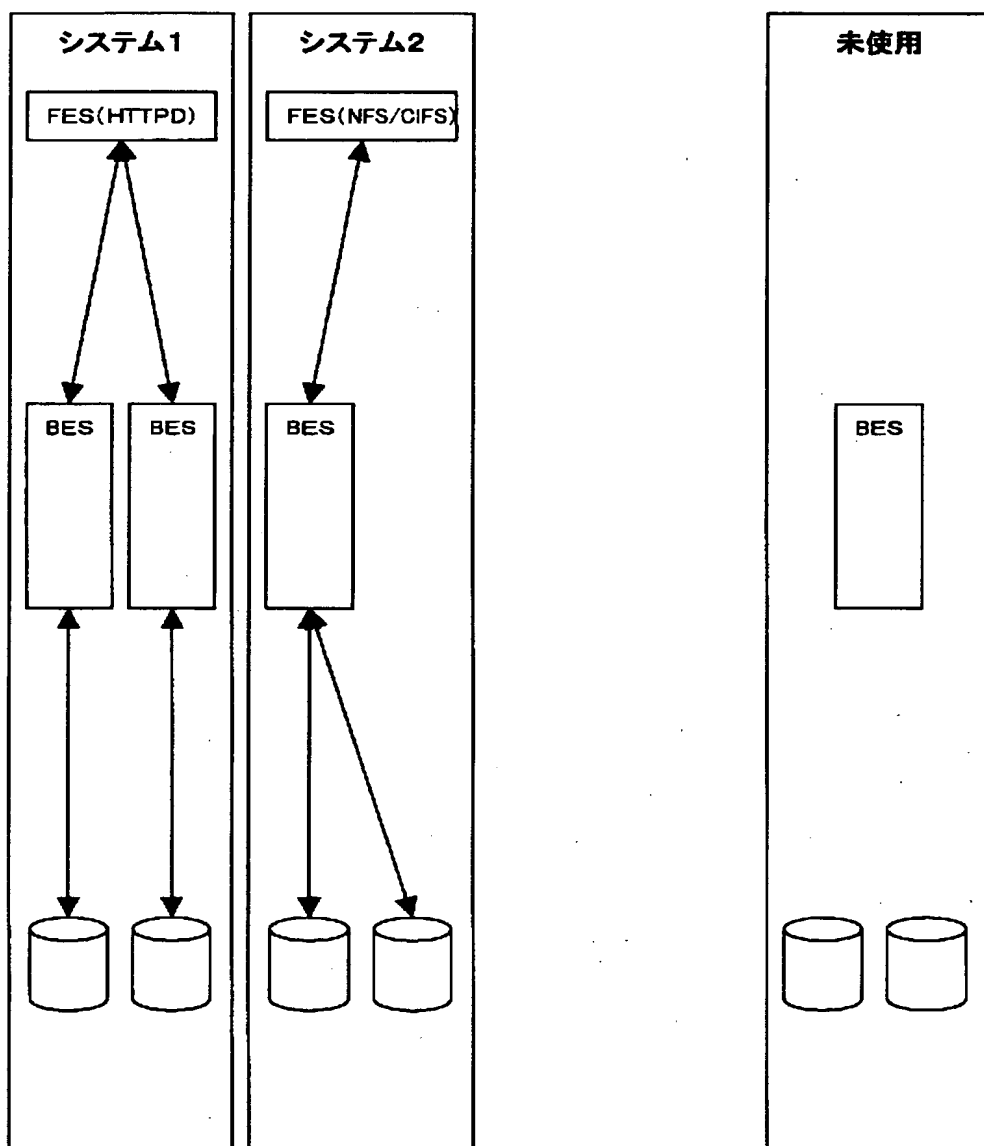
【図 6】

図6



【図 7】

図7



【図 8】

図8

480

BES	所属FES	所属可FESリスト	サーバ負荷
BES1	FES1	FES1	60%
BES2	FES1	FES1	30%
BES3	FES2	FES2	60%
BES4	なし	FES2	0%

FES	システム名
FES1	DBMS
FES2	NFS

【図 9】

図9

485

ディスクID	使用可否	ストレージ装置
DID1	使用	RAID1111
DID2	使用	RAID1111
DID3	使用	RAID1111
DID4	未使用	RAID1111
DID5	未使用	RAID1111

【図 10】

図10

490

ストレージポート名	ストレージ装置ID	ディスク数	ポートI/O数
WWNA	RAID1111	1	3333
WWNB	RAID1111	3	9999
WWNC	RAID1111	0	0

【図 11】

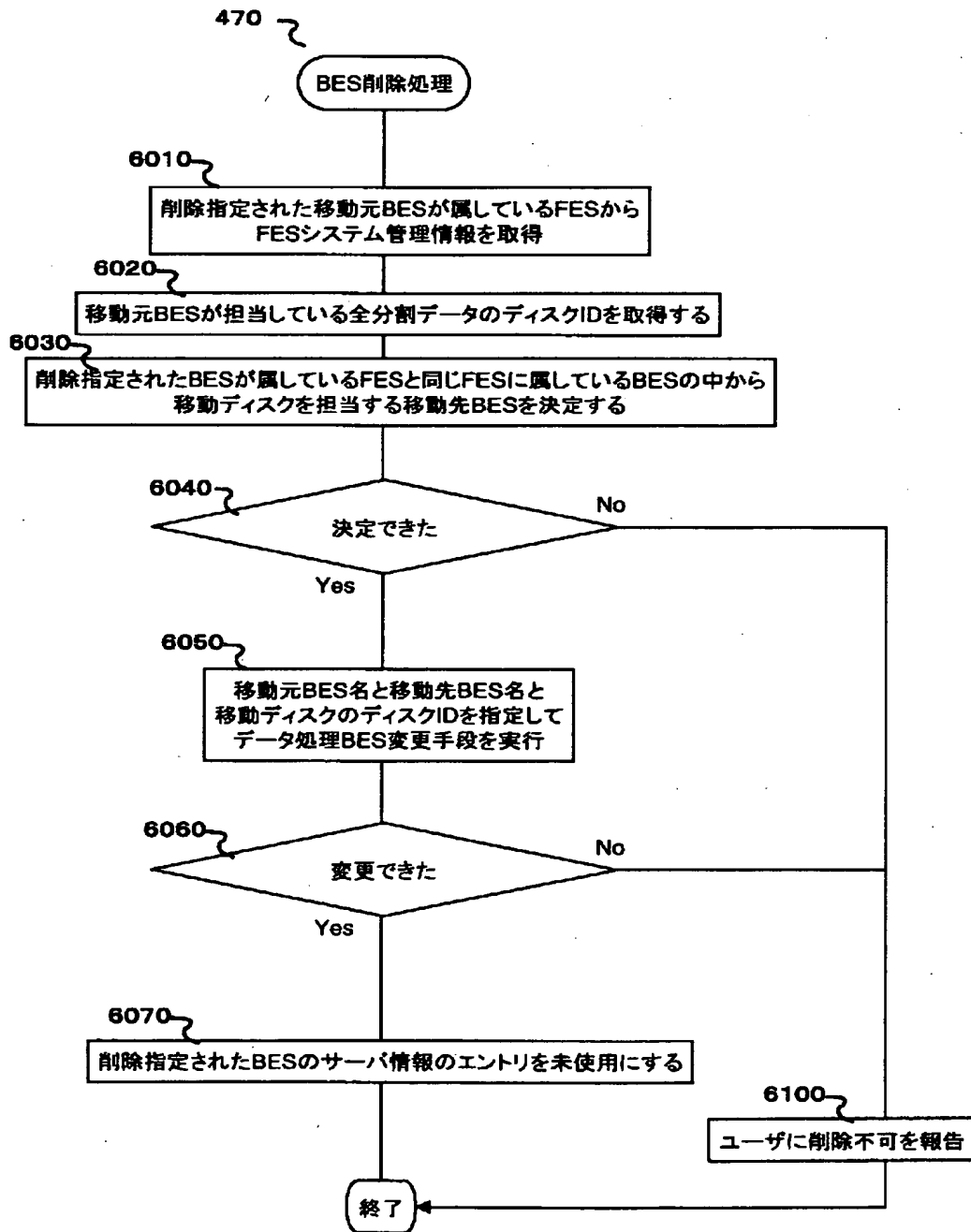
図11

495

サーバ名	サーバポート名	ストレージポート名	ディスクID
BES1	WWN1	WWNA	DID1
BES2	WWN2	WWNB	DID2
BES3	WWN3	WWNB	DID3
BES3	WWN3	WWNB	DID4

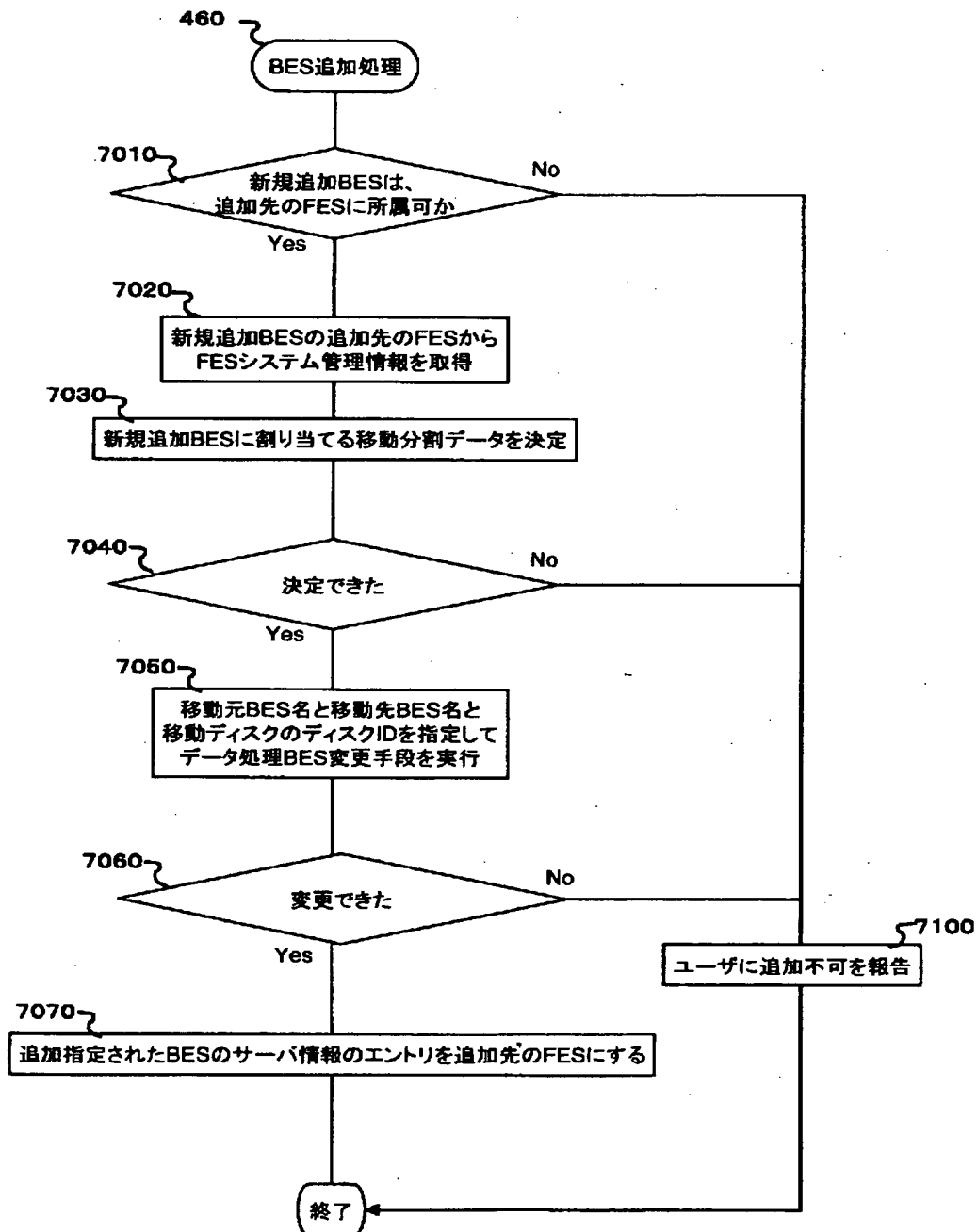
【図 12】

図12



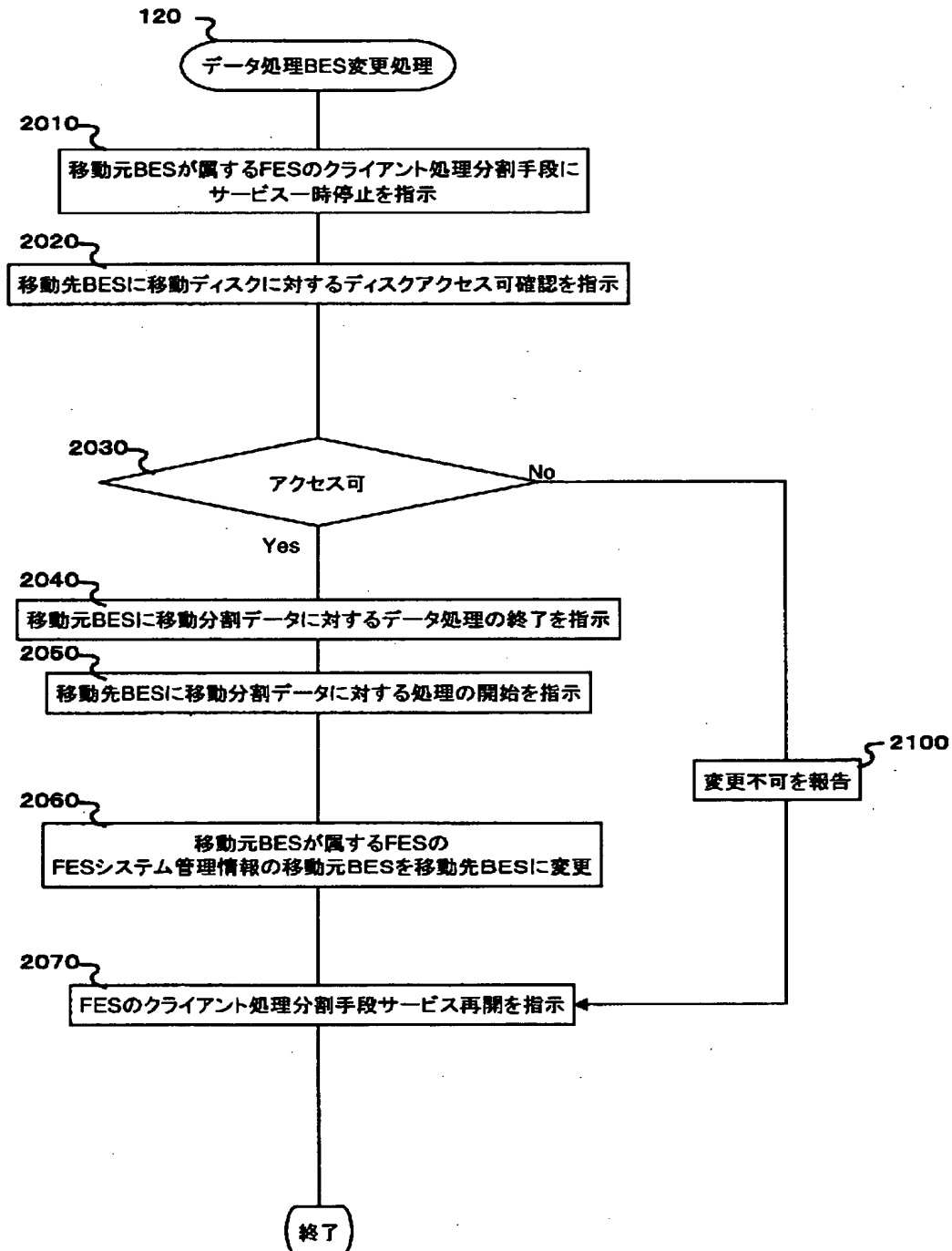
【図13】

図13



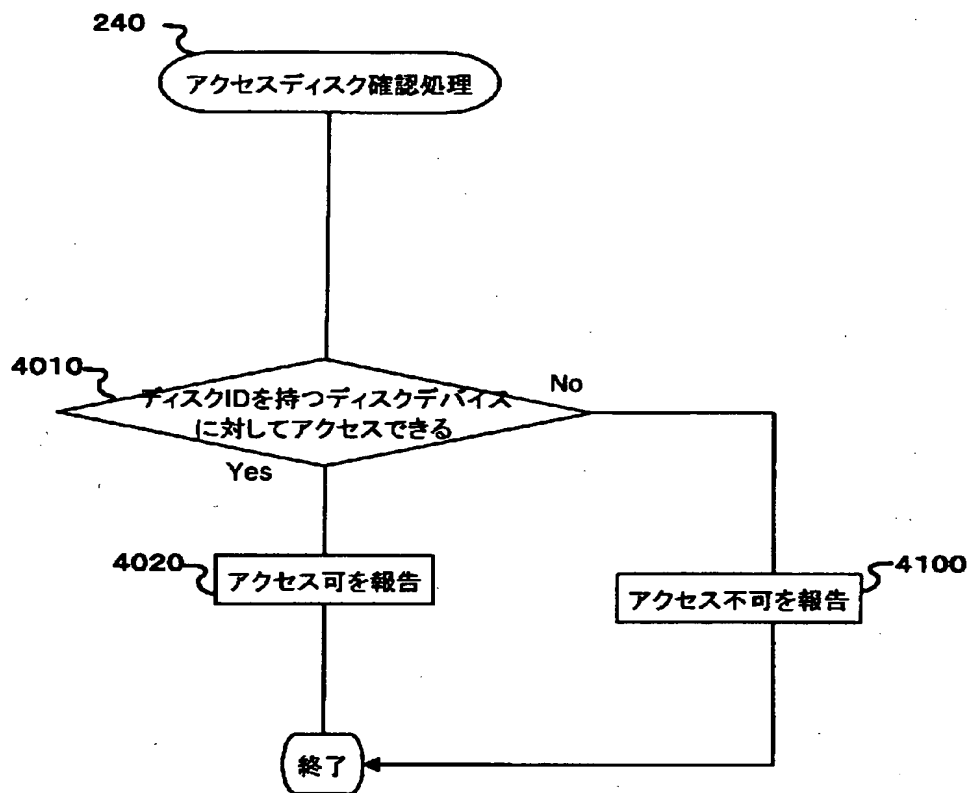
【図14】

図14



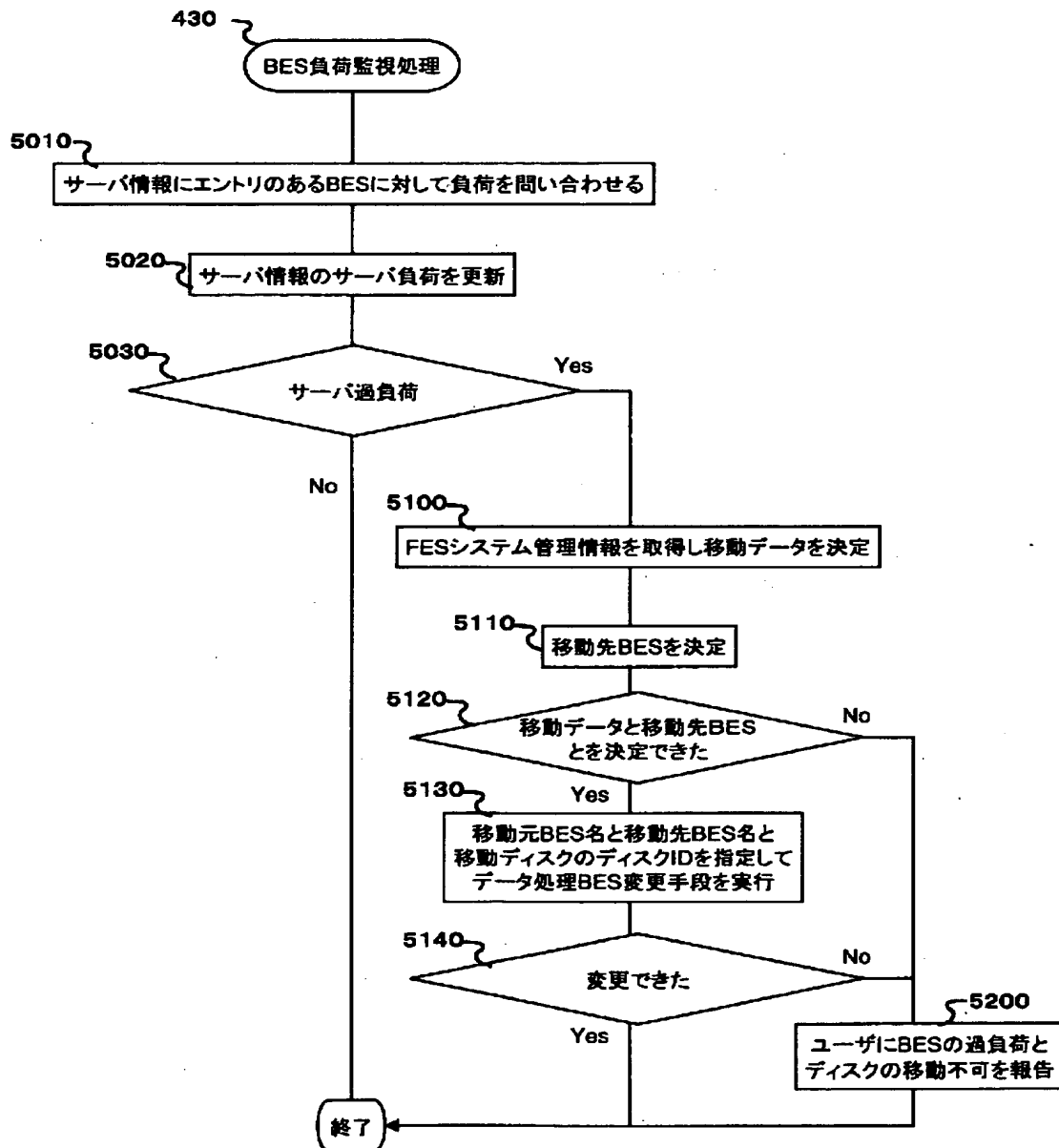
【図 1 5】

図15



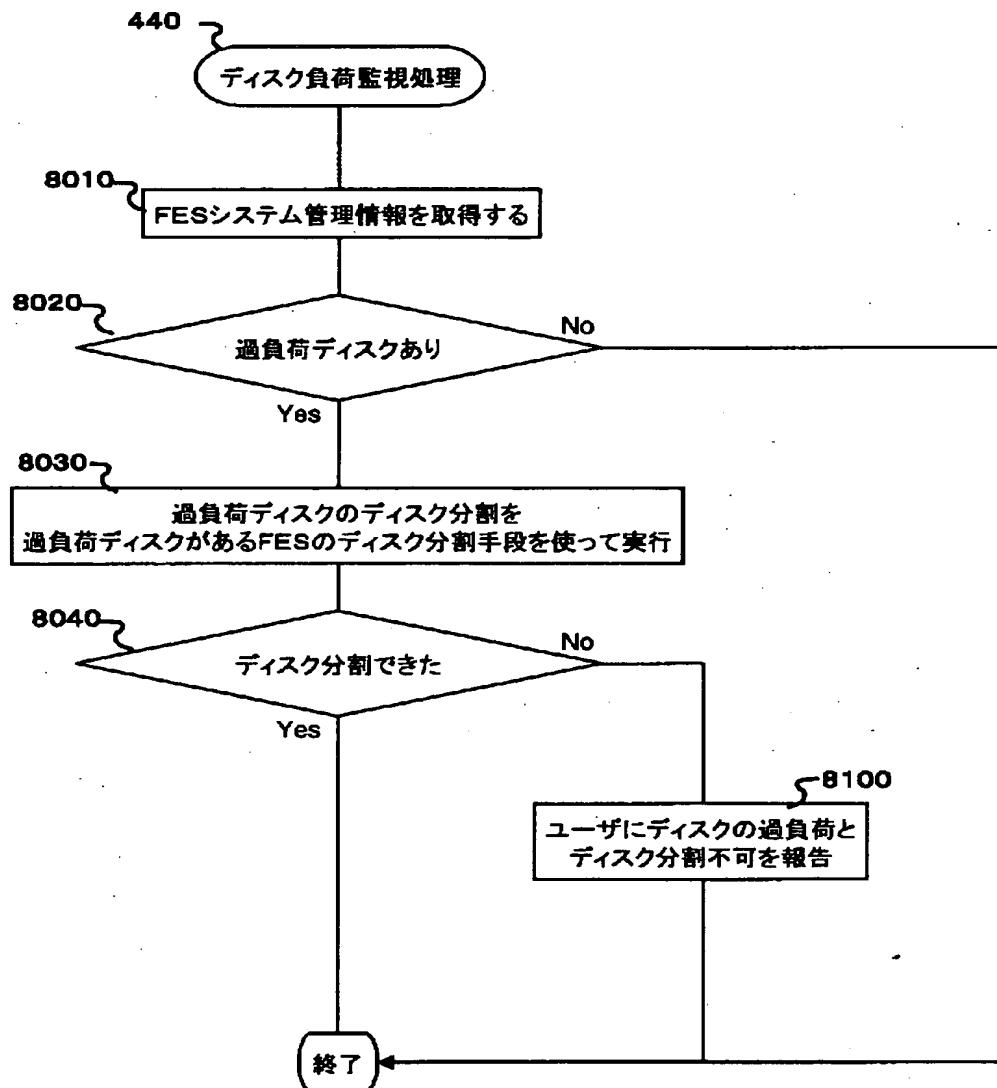
【図 16】

図16



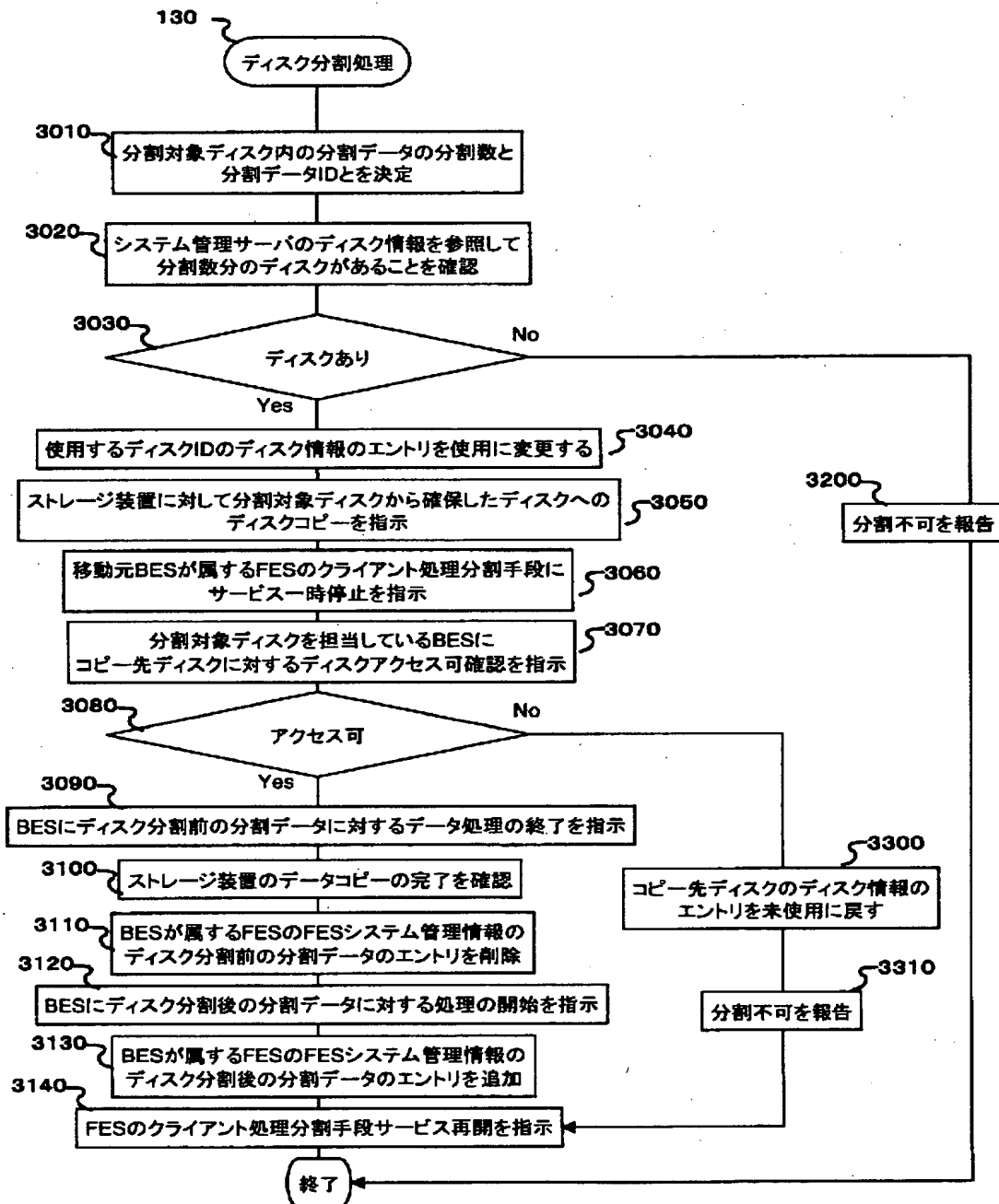
【図 17】

図17



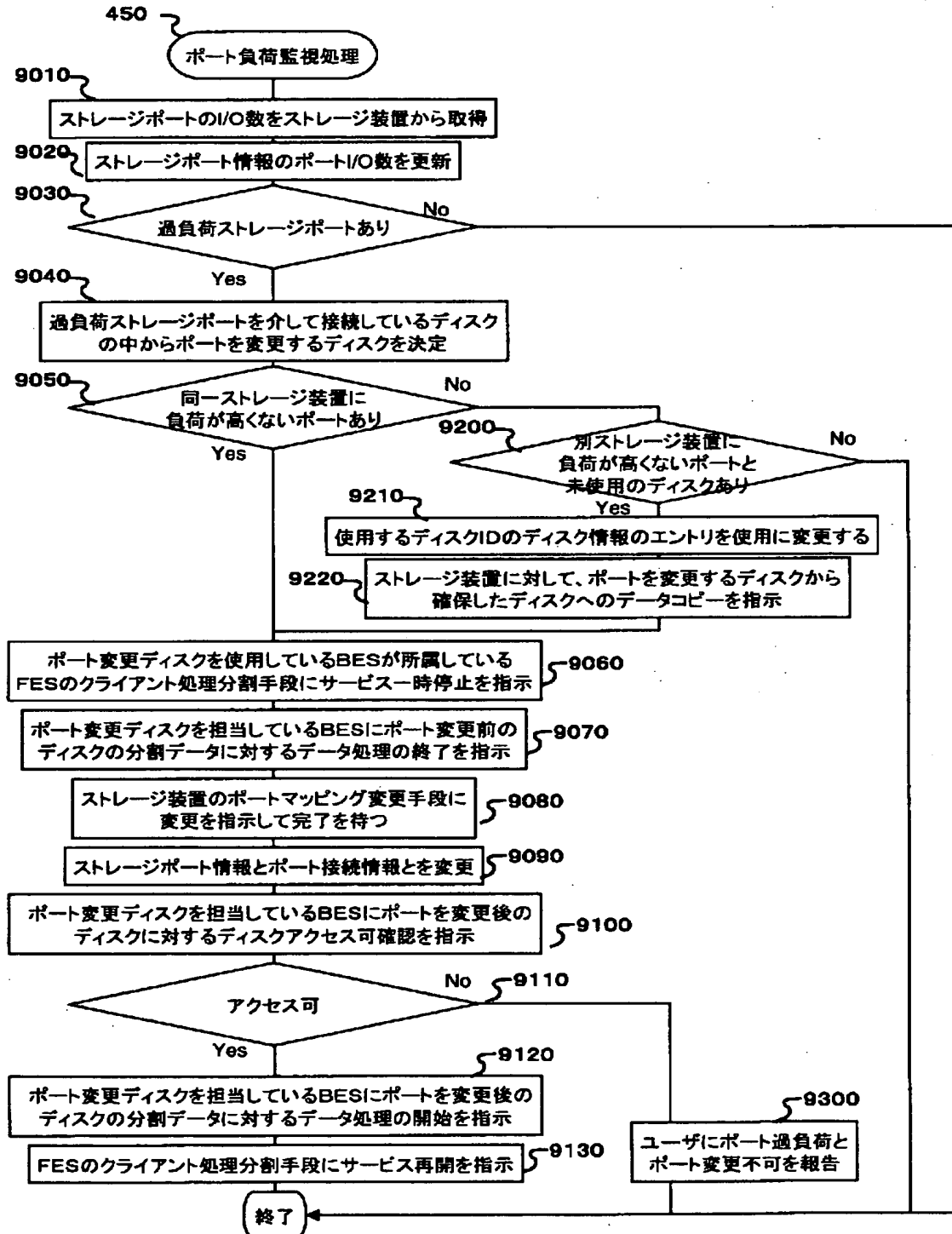
【図18】

図18



【図 19】

図19



【書類名】 要約書

【要約】

【課題】

複数の計算機を使用した分散型の計算機システムにおいて、計算機及びストレージ装置の追加・除去を伴う計算機システムの物理構成の変更を行う場合、計算機システムを停止しての再設定を人手によって行っている。また、ストレージ装置内のディスクと計算機との間の接続があるポートに集中した場合、計算機システムを停止して人手によるポート負荷分散の設定が必要となる。

【解決手段】

ディスクとサーバ計算機をネットワークで接続し、データを格納するディスクに対してアクセスするサーバ計算機を変更して、データを処理するサーバ計算機を変更する。このとき、データを処理するサーバ計算機の変更をシステムが提供するサービスの設定に同時に反映する。

【選択図】 図 1

認定・付加情報

特許出願の番号	特願 2001-179484
受付番号	50100856346
書類名	特許願
担当官	第七担当上席 0096
作成日	平成13年 6月15日

<認定情報・付加情報>

【提出日】	平成13年 6月14日
-------	-------------

出 願 人 履 歴 情 報

識別番号 [000005108]

1. 変更年月日 1990年 8月31日

[変更理由] 新規登録

住 所 東京都千代田区神田駿河台4丁目6番地

氏 名 株式会社日立製作所